DOCUMENT RESUME

ED 183 612 TM 800 020

AUTHOR Stalford, Charles B.
TITLE Educational Evaluation 1979: A View from NIE.
PUB DATE Sep 79
NOTE 28p.: Paper presented at the Annual Meeting of the Evaluation Network (5th, Cincinnati, OH, September 24-25, 1979): Appendix is marginally legible

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Research: *Evaluation: Evaluation Methods: *Evaluation Needs: Federal Government: *Government Role: Government School Relationship: Needs Assessment: Program Evaluation: *Research Committees: Research Needs: *Research Opportunities: Technical Assistance
IDENTIFIERS Cities in Schools: *National Institute of Education: Push for Excellence Program

ABSTRACT

The Testing, Assessment, and Evaluation Division (TAE) of the National Institute of Education was created during a reorganization in 1978. Some of TAE's activities have included: identifying research needs: long-range planning for federally-sponsored evaluation research: reviewing evaluation methodologies: comparing local vs. large national evaluation projects: and monitoring ongoing evaluation research and testing programs. TAE has prepared a planning paper, "Why Study Evaluation?," which discusses three areas of concern: (1) assistance to decision makers to become better consumers of educational evaluations: (2) technical issues concerning evaluation methodology: and (3) the political context of evaluation, which includes the role of the evaluator in program improvement. Current field evaluations focus on two urban improvement programs, Cities in Schools and Push for Excellence. Diverse approaches are likely to be taken in evaluating these programs. ("Why Study Evaluation?" is appended). (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED183612

EDUCATIONAL EVALUATION 1979: A VIEW FROM NIE

A paper prepared for presentation at the annual meeting of the Evaluation Network, Cincinnati, Ohio September 24-25, 1979

Charles B. Stalford
Team Leader for Evaluation Studies
National Institute of Education
Washington D.C. 20208

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Charles B. Stalford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

TM 800020

EDUCATIONAL EVALUATION 1979: A VIEW FROM NIE

In the thoroughgoing reorganization of the National Institute of Education (NIE) that occurred in 1978, there was created a new Testing, Assessment and Evaluation Division. This division incorporated in one body for the first time many of NIE's interests in these areas that had been previously disparate. In total, about $2 million of research to improve evaluation was combined in TAE. While much of TAE's energies have been expended on managing ongoing work, its staff have also been afforded the opportunity for the first time to plan for future research in these areas in wholistic fashion and thus to identify for themselves what they believe to be important.

The papers prepared by Tom Cook and messrs. Hastings, House and Stake as well as my comments are a reflection of this planning process in the area of program evaluation. In the spring of 1979, TAE began its long range planning which had as a focus the fiscal year beginning in October 1980 (in Federal terms, that's fiscal year 1981). The joy of this exercise was that for the moment, at least, we were not obliged to consider budget constraints on our plans; rather, we were to identify areas most in need of further research based solely on their merits. Thus we could and did seek to assess the overall state of program evaluation in education and future needs for research. While the planning exercise was for a somewhat distant date, it nonetheless had the effect for us within TAE of crystallizing and galvanizing our thoughts on what should be done in the nearer term as well.

The results of the planning are reflected in the paper "Why Study Evaluation" produced by TAE in June, which is appended here. In addition to literature reviews and other appropriate activity, we commissioned two papers which were purposefully to reflect different aspects of evaluation research. Tom Cook was asked to write on future developments in quantitative methodologies: Ernest House was asked to write on future developments in nonquantitative and alternative methodologies in evaluation research, including factors affecting utilization. (The Evaluation Consortium at Stanford through Co-chair D.C. Phillips also made available to us a prepublication draft of a forthcoming book on the development and future prospects for evaluation).

While not seeking an exhaustive treatment of evaluation methodologies through the two papers commissioned, we nonetheless obtained papers which describe views at different poles of philosophy on the subject. Indeed it seemed to us that choosing a balance between quantitative and qualitative methodologies to be emphasized in our future research was one major issue for consideration. An emergent interest in our planning was how local evaluations were carried out in school districts versus the sophisticated national and often highly visible evaluations like that of Follow Through, which among other things tend to drive methodological advances in the entire field.

In an effort to learn more about local district evaluations, TAE sponsored a conference of teachers, principals and district research and evaluation directors in the summer of 1978 to discuss needed research in testing and evaluation. In both areas, the participants stressed the need for better knowledge about effective practice in terms comprehensible by local professionals and lay people. For example, means to link testing with instruction to benefit teachers and better means for reporting testing and evaluation results to the public were identified as needs. Concern was not so much about advances in the technology of testing and evaluation which may be commonplace among the participants at a gathering such as this as they were about getting information about effective evaluation practices in use now to practitioners in forms that could be readily understood and used.

When TAE was formed in 1978, it assumed responsibility for monitoring ongoing research programs in evaluation or evaluation-related areas at the Center for the Study of Evaluation (CSE) at UCLA, the Far West Laboratory for Educational Research and Development and the Northwest Regional Education Laboratory (whose program is managed by Nick Smith). In the local area, CSE had in progress a study of local evaluation and decisionmaking which yielded findings strongly similar in many respects to those expressed at NIE's practitioner conference. In particular "evaluation" programs tended to be in fact testing programs: some 75% of respondents to CSE's nationally representative survey of large (over 10,000 enrollment) districts

reported that measurement of achievement is the major purpose behind data collection and testing is the major form of data collection (Lyon et al., 1978). Similar findings have been reported by others (David, 1978; NIE's Compensatory Education Study, 1978). As a consequence, while we in TAE started our planning process with an eye on questions like the relative merits of quantitative and qualitative methodologies, we became increasingly impressed with a need to focus on local evaluations where there may be no "methodology" at all in existence save testing programs.

We have not forgotten the Cook and House papers in our planning, however, which should be evident in our document "Why Study Evaluation?" That document does not discuss our present or planned project activities; the latter could not be published because they are still part of HEW's and the Federal government's as yet unreleased fiscal year 1981 budget. I can give you some idea here, however, of what our present and planned activities may be, which will help clarify our views on some of the many issues so usefully presented to us for consideration by Cook and House.

First of all, the budgetary realities of fiscal 1981 (that is the budget the President will have presented in Jan. 1980 during the election campaign) strongly suggest there will be little new discretionary funding for evaluation or any other area of educational research for that matter. Therefore our emphases will be reflected within ongoing programs and to a limited extent in new ones.

Our paper "Why Study Evaluation?" groups concerns in three areas: assisting decisionmakers to become better consumers of educational evaluations, technical issues and the political context of evaluation. The outcome of our planning process has subsumed the first area of "consumerism" in evaluation under the third heading of the political context; the latter has as one of its explicit goals generation of knowledge which can lead to better utilization of evaluations.

In the area of technical issues, we will continue support for the Research on Evaluation program at NWREL, which is exploring alternative "metaphors" or methods of inquiry for evaluation, such as the journalistic and legal. As the result of an unsolicited proposal, Robert Boruch at Northwestern will do secondary analyses and research to improve methodological techniques over the next three years. I hope some of the quantitative issues raised by Tom Cook will be addressed there. Some of those issues are clearly more susceptible to secondary analysis than others- statistical analysis of data from nonequivalent groups is more feasible than advances in casual modelling in the absence of such consideration by the original investigators- but we will rely on this research at Northwestern to advance technical understanding in evaluation.

In a somewhat different way, our current field evaluations will also contribute to better technical understanding. At the direction of the Secretary's office in HEW, we are evaluating the Cities in Schools and

Reverend Jesse Jackson's Push for Excellence Programs. The former is an ambitious attempt to restructure the delivery of social services in urban areas to focus on needy school children and their families by actually locating those services in schools.

The latter is an attempt to improve urban education in part through a "bootstraps" philosophy of hard work by students. (To quote Reverend Jackson "Its not your aptitude but your attitude that determines your altitude". Both programs pose severe evaluation challenges--they both focus on community, family and institutional structures supporting education rather than classroom and other instructional processes themselves. Push for Excellence poses some unique problems of its own. For one, it is a movement, still in flux as well as discrete programs. For another there is the problem of "attributing" observed effects somehow to the charismatic force of the Reverend himself.

Partially because of the complex nature of the programs, our contractor (AIR) has stressed description of the implementation processes. This is an area of evaluation that in general we do not believe receives sufficient attention. Cooley's reference to the need for better causal modelling of program relationships in advance is related to this point insofar as analysis is concerned. I think there is a more simple point insofar as utilization is concerned. Lacking good description of what

actually occurred in a program, prospective adopters, particularly in schools and communities may not be interested in evaluation results. The most sophisticated (and valid) analyses possible may not lead to utilization unless practitioners can "feel" what happened in a program in their own terms.

Both Cities in Schools and Push for Excellence involve extensive attempts to involve "stakeholders" or those such as parents, community leaders and program officials with a stake in evaluation outcomes in decisions affecting how the evaluations will be conducted. These attempts have been embodied in meetings with stakeholders at the beginning and during the course of the studies, in part to feed back initial findings so that program operators may take them into account in their future plans. The latter aspect of the "stakeholder" process has led to a dilemma of sorts (or intensification of an old dilemma), namely whether the evaluator should intervene to shape the direction of the program being evaluated. For a formative evaluator, the answer, of course, is simple. AIR in these two evaluations has not been commissioned as a formative evaluator, so the issue is not so simple.

Some of you may recall that Robert Perloff's speech as outgoing ERS President last year in Washington, D.C. was explicitly on this subject, reflecting his involvement as a member of the Cities in Schools evaluation technical review panel. In a nutshell, AIR has concluded that some form of intervention by them as evaluators is appropriate from time to time.

We believe that the now-common distinction between "formative" and "summative" evaluation first underscored by Scriven may not be adequate to reflect the work AIR is doing. In particular, one can pass on formative kinds of information in the course of doing a summative study. The Push for Excellence evaluation design is in fact explicitly structured to generate those kinds of information while retaining a central purpose of assessing the effectiveness of the program. In part the flexibility required to perform such an evaluation may be seen as reflecting a need for more diversity in approaches recommended by House et. al. In part also, the stakeholder approach in both of these studies is responsive to the need for fairness in evaluations which House and State are known to espouse. Apart from notions of "justice" and "fairness", we hope that the process will lead to greater utilization of the evaluation results eventually.

Discussion of "stakeholders" has brought us into the second part of our plan, which was identified as studying the political context for evaluations. The stakeholder strategy is in fact one of our activities in this area. Others are in the area of local evaluations which we highlighted previously. It has become increasingly clear to us that many Federal evaluations are of little use to officials and lay people in localities and perhaps at state levels as well. In part the problem is a reflection (but also may reflect) the methodological state of local evaluation found by Lyon et. al.

at CSE: local "evaluation" programs are in reality mostly "testing" programs. In addition reporting requirements for evaluations may frequently favor the Federal or other funding agency sponsoring the evaluation at the expense of providing information which may be useful locally. In a broad sense, this is a political issue, insofar as the "power" of the funding agency determines the shape of the evaluation to be conducted.

The latent "power" of a funding agency to get an evaluation it wants at the expense of local needs may, however, become manifest in part through default. That is, local interests, including research and evaluation personnel, may not have a sufficient bag of methodological "tricks" appropriate to meet their own needs. Such needs might be efficient ways to accomplish classroom observation of programs, knowledge about nonexperimental research procedures that would yield results responsive to local needs (including better program descriptions) and better ways to disaggregate and report evaluation data locally.

TAE has taken two new steps to address local evalaution needs in addition to activities already underway mentioned earlier. In conjunction with USOE and other divisions of NIE, a study has been initiated to study uses of testing and evaluation information at state and local levels. Buinding in part on earlier CSE work, the study will investigate districts where such information is being used to determine influences producing the use. In addition, better strategies for providing technical assistance locally will be explored. This study will continue for two years.

The second step is that the fiscal 1980 grants announcement for NIE's Teaching Learning Program (in which TAE is lodged) will include a "Methodology" section dealing with both testing and evaluation. The evaluation section will be explicitly addressed to school district evaluation needs. Districts are invited to submit grant proposals with or without outside support (including state education agencies) to perform research leading to better evaluation methodology suitable at the local level. It is our intent to give preference to evaluation proposals coming from local districts, which is something of a novel precedent in such an announcement.

I anticipate we will pay increasing attention to local and state evaluation needs in our future work to the extent our resources permit. To come full circle, our intent in part is to bring some of the sophisticated methodological thinking evidenced in Tom Cook's paper to bear in appropriate form on local and state needs. At the same time, we will rely on research at Northwestern, NWREL and our own field evaluations to move the state of the art ahead. (We have recently assumed responsibility for monitoring Bill Cooley's research on evaluation program at the Learning Research and Development Center also, which should add to our strength in this area.)

Our work on understanding the political context of evaluation will take part in our own field evaluations. (Because we don't trust ourselves to assess the results of those efforts entirely, we have recently initiated a new study which will in part evaluate our stakeholder strategy in the

Cities in Schools and Push for Excellence studies). In part, it will be based on our efforts to improve the methodology and utilization of evaluations at local and state levels. While the local and state orientation of our research program is still emerging, I for one am inclined increasingly to describe that as a main thrust of TAE's program in its own right. I would guess at this point that diversity of approaches to evaluation, as suggested by House et al. will increasingly characterize evaluations at these levels. If it doesn't, it should.

REFERENCES

David, Jane L. Local Uses of Title I Evaluations. Menlo Park: SRI International. July 1978 (Research Report EPRC 21).

Lyon, C.D., Doscher, L., McGranahan, P., & Williams, R. Evaluation and School Districts. Los Angeles: University of California, Center for the Study of Evaluation, 1978.

National Institute of Education. Compensatory Education Study: Final Report to Congress. 1978.

(Attachment: NIE planning paper "Why Study Evaluation?")

ATTACHMENT

RESEARCH AREA PLAN: EVALUATION

1. Why Study Evaluation?

The Federal government got into evaluation in a big way in 1965, with passage of the Elementary and Secondary Education Act. Into this Act, Senator Robert F. Kennedy wrote provisions intended to give greater power to parents and community members, including the power information provides through evaluation. Prior to 1965, Federal evaluation requirements had focused more on accountability: were funds reaching the beneficiaries intended by Congress, were they being spent on the services intended, did they lever the additional resources which were a major purpose prior to 1965 of Federal involvement in education? After 1965, a new set of purposes for evaluation was added: were the Federal funds achieving the goals intended for learners, including academic achievement, educational attainment, and equity?

Experience since 1965 has led to something of a paradox. First, in many instances evaluations which have been severely criticized on methodological grounds have been used to make policy, to affect practice and most of all, to develop a climate of belief about what is possible for educational change. One example is the Performance Contracting Experiment, which by-and-large wiped out any enthusiasm for this approach to educational change. Another is the Westinghouse Study which halted large investment in early childhood education dead in its tracks for years. A third is the Coleman study, which influenced a climate of belief about the effects of differences between schools relative to home differences in educational achievement. Almost all of these studies have had "negative" effects, if these are defined as contributing to belief that education is relatively powerless and can't be changed much anyhow, even if we knew better what worked.

The persons most affected by these evaluations---because they are reported primarily for compensatory education programs---are the poor, the minorities, and the least privileged in our society. Programs on their behalf are by and large getting shot down by evaluations, particularly large scale national studies. The General Accounting Office, in a September 8, 1977 report to Congress (Problems and Needed Improvements in Evaluating Office of Education Programs) has discussed a moratorium on the large national studies.

Second, the other side of the paradox is that evaluations are written into almost all educational legislation as a means for local program improvement, and into demonstration programs as a means of learning from experience lessons of value to national policy and practice. Congress has become highly prescriptive in this legislation, often in response to advice from evaluators, and has included such

features as: (1) prescribing the criteria by which ALL programs will be measured ANNUALLY at the local level (Vocational Education Amendments of 1976, specifying placement in training related jobs, employer satisfaction and continuation in data to be collected from all programs) and (2) through the Education amendments of 1978, prescribing the use at local and state levels of evaluation data for local program improvement and planning. Elsewhere, Congress has required that certain evaluation designs be used, that technical assistance be available to all LEAs and SEAs in implementing these designs, and has mandated nationally aggregated data on program implementation and effectiveness.

The paradox is that evaluations on the one hand are said "not to be utilized" as our greatest problem, but they in fact have been utilized nationally with macro-negative effects on belief about what schooling can accomplish and the educability of minority and low income children; on the other hand, they are said to suffer too many technical defects to be usable, yet through legislation and regulations, the country is going hell-for-leather to implement them in every hamlet, village, town and city which receives Federal funds. Even without the Federal enthusiasm, states and local agencies have institutionalized evaluation programs, and to the Federal requirements have been added state requirements and school board expectations for accountability and evaluative study.

Again, most of these studies affect the poor and minorities, and many use as outcomes measures of basic skills and academic achievement, areas of great concern to the Teaching and Learning Program. In particular, four areas for educational evaluations are likely to be of considerable importance in the years ahead: bilingual education the new basic skills legislation, adult basic education and education for the handicapped. Seen more broadly, there will probably be continuing need for evaluative evidence regarding the effects of such policies as busing to achieve school desegregation, the effects of various practices implementing Title IX (Women's Educational Equity), resource concentration by school closings, and mainstreaming as an approach to serving handicapped children.

To the extent that T&L and NIE have a stake in contributing to wiser, better informed decisions about educational practices and policies, we need to do what we can in three areas relating to evaluation:

1. Finding out how to help educational decision-makers at all levels become wiser consumers of educational evaluations, particularly those affecting bilingual education, basic skills, adult basic education and the handicapped.

2. Examining the technical issues each of these areas presents in developing adequate evaluations: beyond measurement of process and

outcomes (if we only knew the right variables and could measure them), what else technically in designs, analyses and interpretation will present unique problems in these programs, and how can we best deal with these technical problems?

3. Studying the political context in which evaluations are embedded that affect wise identification of the questions to be asked, the way in which the evaluations are designed and conducted, and their eventual utilization- with particular reference to unique political factors that will affect utilization of evaluations on bilingual education, on basic skills, on adult basic education, and on the handicapped.

2. The State of Practice and Evaluative Research

A. Finding out how to help educational decision-makers at all levels become wiser consumers of educational evaluations

Considerable effort has gone into different approaches to helping educational consumers of evaluations become wiser. These approaches include:

- o attention to the information provided (e.g., Brickell et al., 1974)
- o experiments in non-traditional presentation of results e.g., "adversial" or "jury-trial" models of presenting findings, which use the debate format for the decision-makers (Wolf, 1975)
- o development of guides to those who commission or purchase educational evaluations, checklists of what every project director should know before signing a contract for an evaluation or approving an evaluation plan (e.g., Educational Testing Service, 1977)
- o a plethora of training seminars for school board members, superintendents, principals and teachers offering crash courses in experimental design, evaluation planning, setting objectives, summative and formative evaluations and other topics in vogue
- o some provision of pre-service training in evaluation, although very few schools of education provide either masters or doctoral level programs in educational evaluation
- o development by the profession since 1974 of evaluation standards (e.g., Stufflebeam et al. effort funded by NIE

and NSF) which presumably would be helpful to evaluators and consumers

- o extensive articles on evaluation in most subject area journals concerned with education.

The extent to which these and other efforts are making a contribution to wise use of evaluation findings is unknown. In fact, very little is known about the uses of evaluation at the local level. The available evidence (GAO, 1977; CSE, 1978, David, 1978) suggests evaluations are not being used for the constructive purposes of program improvement they are presumably intended for. Much of our present knowledge about evaluation use is anecdotal, and concern for helping decision-makers become better consumers stems more from incidents than well-documented instances of wise use or of abuse. These incidents do suggest, however, reasonable concern:

- o the state of California has pioneered an early childhood education program. The evaluation of this program has been severly critiqued by some experts and defended by others. Much of the critique hinges on the research design: the comparability of the participants and non-participants, and the lack of knowledge of their probable developmental "curves", absent the program

- o reviews of local evaluations in the ERIC system have suggested that practice is lagging far behind what already is known. For example, covariance techniques are still applied as statistical band-aids when the conditions of their appropriate use are wholly absent and effect of their use problemmatic (Cronbach et al. 1978)

- o few evaluations are relating process and outcome measures- the program descriptive data tend to be reported in one section, and the outcome data in another, without application of the paradigms long available (e.g., Stake, 1967; Stufflebeam, 1971) to relate program content and implementation to its outcomes.

- o at the national level, big evaluations still are being uncritically used to form national policy and belief (e.g., the 1974 Rand report on change agents; the 1978 AIR report on bilingual education), suggesting a considerable need to improve the sophistication of evaluation consumers.

In the four programs proposed for focus, local, state and national-level evaluations have been required by Congress. In addition, numerous discrete demonstration programs have been funded. Consumers by the hundreds will be grappling with evaluation findings in basic skills,

bilingual education, adult basic education and education for the handicapped. When over 70% of classroom teachers can't correctly interprete a grade equivalent score or a percentile, the probable level of naivete on more complex aspects of evaluation may take on something of a broad national concern in the four programs mentioned.

B. Examining the technical issues each of the areas presents

Evaluation research has tended to be general. There now are many broad technical documents on matters such as ways of identifying the objectives held by various stakeholders e.g., the Multi-Attribute Utility model; (Edward Guttentag and Snapper, 1975) the Context, Input, Process, Product model-CIPP (Stufflebeam, 1971); on alternative algebras for coping with non-equivalent control and comparison groups (Riecken and Boruch, 1974; Boruch, in Boruch and Riecken, 1975; Cook and Campbell, 1976, Cook, 1979); and on integrating qualitative and quantitative findings in evaluation (Boruch, 1975; Tikunoff and Ward, 1977.)

Each of these areas needs further exploration:

- As Scriven has point out (1978) the so-called need analyses typically are merely "wants" analyses; he regards the state of the art in this area as primitive and lamentable

- Relating what is known about important factors influencing acquisition and use of reading, writing, language and arithmetic skills and the ways in which these skills develop in ways which are evaluable is equally primitive. Evaluation is inherently a study of change, and very little is known about change and development--the expected growth of children or adults on relevant measures. Our research is basically short-term and cross-sectional: our need is basically longer-term and longitudinal in almost every aspect of development to be studied in education.

- Relating what is known about development to evaluation designs is virtually non-existent. Only a few (Cooley, 1976; Magidson and Abt, 1979) draw attention to the need for conceptual models of developmental processes as integral parts of evaluation designs. Times of pre/post measures are usually chosen to coincide with convenience--school years--rather than rates of growth. This would be just dandy if growth were linear in all circumstances for all persons for all possible outcomes, but it probably is not.

- The ways of coping with a statistics based on very demanding assumptions about random distribution of variations with a

real world that does not often permit evaluations to be designed so that the statistics can justifably be used continues to be an enormous technical barrier. While some techniques of converging analytic models on the same set of data, secondary analysis (Bryant and Wortman, 1978; Boruch, 1978) or pooling findings across studies in meta-evaluation(Glass, 1976) seem more promising than continuing to bash away de novo at design issues, others (e.g., Haney and Weisburg, 1978) are pursuing new development models which may in fact lead to feasible field designs. There simply is not adequate basic information on how far off conclusions will be if we play fast and loose with the statistical assumptions in either quasi-experimental designs or statistical fancy footwork after the data are in. Further research in this area is needed.

It can be argued that these technical issues should be pursued without the possibly added burden of special considerations associated with subject areas such as reading, writing, mathematics, or bilingualism. The evaluation research community may be well-started, however, on the broad technical questions without further assistance from NIE. There are many other sources of funds in the government for pursuing these broad questions, and some get addressed as a by-product of national and other evaluations.

What is missing is putting these subject matter considerations together with evaluations BEFORE an RFP has to go out in two weeks to identify unique technical issues and ways of addressing those. As illustrative examples:

- o in an evaluation of a writing program, how important will it be to obtain contextual information about the non-school demands for writing and opportunities for rehearsal in the home? How much difference would this expensive, hard-to-get information make in conclusions affecting evaluations of in-school program effectiveness?

- o in an evaluation of a bilingual education program, what is known about rates of acquisition of first language fluency among monolinguals in the various sub-skills of oral and written language use that ought to affect when data are collected: is there usually, for example, a rapid point-of-inflation at age three, then again at eight? If so, and the years in between are typically plateaus by most measures, an evaluation measuring change in six months (October to April) at age six years might be highly inappropriate.

- o in evaluations of mainstreaming handicapped children, what is known about teachers' training in handling classroom diversity, in needs of handicapped children, and in the attitudes of other children toward handicapped peers that should be process variables, and what analytic techniques (case studies? blocking? regression?) best fit the nature of the data that ca

be collected on these process variables?

The need for integrating subject matter with technical considerations is currently unfilled in educational evaluation research. Research of this kind is distinguished from designing a specific evaluation by (a) basing the analysis of the unique demands on evaluations in a broad subject area theory (subject matter experts' participation) as well as in evaluation technology, (b) developing a plan for evaluative research well enough ahead of time to provide a chance for getting useful information in contrast to seeking designs under the pressure of getting out a specific study, and (c) linking the evaluative research to practice in education in a subject area. This research approach would not preclude in any way learning from evaluation technology in other disciplines (e.g., of community mental health centers, of food and nutrition programs, etc.) or eventual generalization to other evaluation methodologies, but it could provide links now missing between the existent evaluation technologies and studies of specific educational content.

C. Studying the political context in which evaluations are embedded

According to many(Weiss, 1973; Abt, 1976; Freeman, 1977; Cronbach, 1978; Stake and House, 1979)some of the biggest problems in evaluation have little to do with evaluation technology, and a great deal to do with the political context of evaluation. There are at least two facets to this context.

First, examining the trail of Federal legislation, Federal regulations, state legislation, state regulations, and local evaluation requirements makes clear that there are multiple reasons why evaluations are done at all. These include the belief that evaluations are necessary for minimum public accountability at a time when a certain fiscal conservatism, of emphasis on better use of public funds rather than program expansion, and the what-are-we-getting-for-the-money issues is prominent. A second reason according to this trail, is the hope that better information will lead to better program decisions and to program improvement. This purpose has longer roots (e.g., Cronbach, 1963) but continuing vitality. A third reason is political enpowerment. Lasch (1978) has argued that technotwits have expropriated political power through high technology, including information presented in inaccessible forms and argues that redistribution of power to citizens requires equal access to such information. A fourth reason is goosing the system: the belief that folks who look over their shoulders to see who is watching or gaining on them run faster.

Second, much has been written recently about education as a loosely coupled system, and great attention has been given to the organizational characteristics of educational institutions: who reports to whom, influences on different individuals and ways in which their decisions and power do and do not cross the action synapses to other parts of the

loosely coupled system. CSE(1978)has preliminary data showing that organizational factors significantly affect local evaluation usage. One response to this emerging belief in evaluation has been attention to stakeholder involvement: if decisions are influenced by many people, any one of whom can block transmission into action, then stakeholders must be involved in all aspects of evaluation(Smith, 1976; Weiner et el., 1977).

Very little is known empirically about these contexts. Research on utilization of evaluations tends to focus on the big national policy studies (e.g., Williams, 1971; Weiss, 1977; Young, 1978)and on the political forces which have blocked their use e.g., Steiner (1976) on children's policy. There is considerable belief that evaluations are not utilized and that the primary reason is not technology, but lack of stakeholder involvement and attention to the way in which change comes about---if at all---in a loosely coupled education system.

Much of the research is in the form of case studies and each new generation of evaluations tends to incorporate accumulating beliefs about the structural and political factors influencing the utilization of evaluation. As an example, the NIE Cities in Schools and PUSH-EXCEL evaluations use the Multiple-Attribute Utility model and stakeholder involvement more systematically and thoroughly than any prior study of this magnitude (except perhaps the Edwards report on school desegregation to the city of Los Angeles). If the evaluation results are used, one interpretation could be that the stakeholder involvement assumption is justified.

There are some instances where evaluations have been utilized at a local level to improve practice. CSE has noted individual school districts where evaluations are a major source of information affecting improvement and allocation decisions. Alkin et al. (1974) found that evaluation can be used both formatively and summatively even in a mandated (Title VII) situation. But these reports are few and far between. Relatively little attention is directed to use of evaluations on the local level by practitioners. TAE's summer 1978 practitioner conference on testing and evaluation in schools identified several areas for further research on local use of evaluations including means to reduce Federal and state regulatory burdens, facilitate cooperation between evaluators and teachers and to present evaluation findings more effectively (Kean, 1978; Holley, 1978).

If the hunches of Cronbach, Freeman, Weiss, and others are correct---and evaluation is a highly politically charged activity--- we can expect the charge will be great in the four areas chosen

for examination: bilingual education (where lightning already plays furiously about the lightning rod of the AIR study), education for basic skills, adult basic education and education for the handicapped. These areas are fruitful ground for systematic exploration of the influence of the political context on evaluation use and what can be done to work with these political flows.

As an example, the stakeholder group models tend to assume stakeholder interests are defined by status (teacher, principal, parents, director of instruction); that the interests are relatively legitimated by these positions (President of the local teachers union, Chairperson of the school board, chief state school officer) etc., and that the interests within a group (parents; teachers) are much more convergent than interests between groups. In bilingual education, this may not be so. The significant splits in stakeholder interests may break along quite different lines, associated more with beliefs about transition, maintenance, or cultural heritage aspects of the programs. How to identify <u>what</u> interests are significant; to identify representatives of these interests who will have legitimacy with the stakeholders when these do not follow status lines; and how to deal with the issue of whose story to tell (e.g., Cohen and Farrar, 1978), when funds and time are limited require considerable methodological research on the political aspects of evaluation, broadly defined.

REFERENCES

Abt, Clark C., The Evaluation of Social Programs. Beverly Hills: SAGE Publications, Inc. 1976.

Alkin, Marvin C., Koscekoff, J., Fitzgibbon, C., and Seligman, R. Evaluation and Decisionmaking: The Title VII Experience. CSE Monograph Series in Evaluation, No.4. Los Angeles: Center for the Study of Evaluation, UCLA. 1974.

American Institutes for Research. Evaluation of the Impact of ESEA Title VII Spanish/ English Bilingual Education Programs. USOE contract OEC-0-74-9331. Palo Alto. 1978.

Boruch, Robert F. "On Common Contentions About Randomized Experiments for Evaluating Social Programs" in Experimental Testing of Public Policy. Boruch and Riecken (eds.). Boulder: Westview. 1975.

Boruch, Robert F.(ed.) Secondary Analysis. An edition of New Directions in Program Evaluation, 1978, 1, 4.

Brickell, Mitchell M. et al. Data for Decsions: An Analysis of Evaluative Data Needed by Decisionmakers in Educational Programs. New York: Educational Research Council of America. 1974.

Bryant, Fred B. and Wortman, Paul M. "Secondary Analysis: The Case for Data Archives". American Psychologist, 1978, 33, 4, 381-387.

Cook, Thomas D. "Working Paper on the Methodological State of the Art in Quantitative Evaluation Research, with Tentative Directions About New Directions to Follow". Working paper prepared for Evaluation Team planning process, 1979.

Cooley, William W. "Explanatory Observational Studies". 1978 AERA Annual Meeting Division D State of the Art Address. In Educational Researcher, 1978, 7.(9), 9-15.

Cronbach, Lee J. "Course Improvement Through Evaluation". Teachers College Record. 1963, 64, 672-683.

Cronbach, Lee, J. et al."Analysis of Covariance in Nonrandomized Experiments: Parameters Affecting Bias". Occasional Paper from the Stanford Evaluation Consortium. Stanford: 1978.

Cronbach, Lee. J. "Designing Educational Evaluations". Occasional Papers from the Stanford Evaluation Consortium. Stanford: 1978.

CSE Interim Report "Evaluation and Decisionmaking in School Districts" (Catherine Lyon, Project Director). Los Angeles: Center for the Study of Evaluation, UCLA, 1978.

David, Jane L. Local Uses of Title I Evaluations. Menlo Park: SRI International. July 1978 (Research Report EPRC 21).

Educational Testing Service. A Guide to Evaluation Training Materials. Princeton. 1977.

Edwards, Ward; Guttentag, Marcia and Snapper, Kurt. A Decision-Theoretic Approach to Evaluation Research" in Handbook of Evaluation Research, Vol 1. Struening and Guttentag eds. Beverly Hills: SAGE Publications, Inc. 1975.

Farrar, Eleanor; DeSanctis, J. and Cohen, David. "Alternative conceptions of Implementation". Draft report to NIE pursuant to contract 400-76-0163. Cambridge: Huron Institute. October, 1978.

Freeman, H.,"The Present Status of Evaluation Research" in Evaluation Studies Review Annual, Volume 2. Guttentag and Saar, ed. Beverly Hills: SAGE Publications, Inc. 1977

General Accounting Office. Problems and Needed Improvements in Evaluating Office of Education Programs. Washington D.C. 1977.

Glass, Gene. "Primary, Secondary and Meta-Analysis of Research". Educational Researcher, 1976, 5, 10, 3-8.

Holley, Freda M. "What it Takes to Win". Prepared for TAE Summer 1978 conference of practitioners to discuss needs for research on testing and evaluation in schools (unpublished).

House Ernest R. and Stake, Robert E. " A View of Program Evaluation in 1979".Working paper prepared for Evaluation Team planning process.

Kean, Michael H. (Executive Director, Office of Research and Evaluation, Philadelphia Schools) " Responding to Conflicting Evaluation Demands". Prepared for Summer 1978 conference of practitioners.

Lasch, Christopher. Haven in a Heartless World: The Family Besieged. New York: Basic Books, Inc. 1977

Magidson, Jay and Abt, Wendy. Quixotic Pupil Effects. Cambridge: Abt Associates. 1979 (in preparation).

National Institute of Education. Compensatory Education Study: Final Report to Congress. 1978.

Rand Corporation. Federal Programs Supporting Educational Change. Prepared for USOE, HEW contract OS-73-216. Santa Monica. 1974.

Riecken, Henry W. and Boruch, Robert F. (ed.) Social Experimentation. New York: Academic Press. 1974.

Rivlin, Alice M. Systemmatic Thinking for Social Action. Washington D.C.: Brookings Institution, 1971.

Scriven, Michael and Roth, Jane. "Needs Assessment: Concept and Practice". New Directions for Program Evaluation, 1978, 1, 1.

Smith, Stephen. The Utilization of Evaluative Information. Washington D.C.: Team Associates, Inc., 1976.

Stake, Robert E. "The Countenance of Educational Evaluation". Teachers College Record, 1967, 68, 523-540.

Steiner, Gilbert Y. The Children's Cause. Washington D.C.: Brookings Institution. 1976.

Stufflebeam, Daniel L. et al. Educational Evaluation and Decision-making. Itasca, Illinois: F.E. Peacock Publishers, Inc. 1971.

Tikunoff, William J. and Ward, Beatrice. Special Issue: "Exploring Qualitative and Quantitative Research Methodology in Education". J. Anthropology and Education, May 1977, VIII.

Weiner, Stephen S. et al. "Pathological Structures in Institutional Evaluations and a Possible Cure" (unpublished). Stanford Evaluation Consortium. 1977.

Weiss, Carol H. (ed.) Using Social Research in Public Policy Making. Lexington, Mass.:Lexington Books, 1977.

Williams, Walter. Social Policy Research and Analysis: The Experience in the Federal Social Agencies. New York: American Enterprise Elsevier Publishing Co. 1971.

Wolf, R.L. "Trial by Jury": A New Evaluation Method, 1." Phi Delta Kappan, November 1975.

Young, Carlotta J. "Evaluation Utilization". Paper presented at Evaluation Research Society Second Annual Meeting, Washington D.C., November 2-4, 1978.